

Claims

1. Method of animating a synthesised model of a human face driven by an audio signal, comprising an analytic phase, in which an alphabet of visemes is determined, i.e. a set of information representing the shape of a face of a speaker corresponding to phonetic units extracted from a set of audio training signals, and a synthesis phase, in which the audio driving signal is converted into a sequence of phonetic units associated to respective temporal information, whereas the sequence of visemes, corresponding to the phonetic units of the set comprised in the audio driving signal, are determined in the analytic phase, and the transforms required to reproduce the sequence of visemes are applied to the model

characterised by the fact said analytic phase provides an alphabet of visemes, determined as active shape model parameter vectors, to which the respective transforms of the model, expressed as parameters of low-level facial animation compliant with standard ISO/IEC 14496, are associated. During both the analytic phase and the synthesis phase, the sequences of visemes, corresponding to the phonetic units of the audio training signal and of the audio animation driving signal, respectively, are transformed into continuous representations of movement by means of viseme interpolation, conducted as convex combinations of the visemes themselves to which combination coefficients, which are continuous functions of time, are associated, the combination coefficients carried out in the synthesis phase being the same as those used for the analytic phase combination.

2. Method according to claim 1 characterised by the fact that the coefficients of said convex combinations are functions of the following type:

$$\beta_n(t) = \begin{cases} \cos^2\left(\frac{\pi}{2} \frac{t-t_n}{t_{n+1}-t_n}\right); & t \in [t_n, t_{n+1}] \\ \cos^2\left(\frac{\pi}{2} \frac{t-t_n}{t_n-t_{n-1}}\right); & t \in [t_{n-1}, t_n] \\ 0; & t \in [t_{n-1}, t_{n+1}] \end{cases}$$

- a.
3. Method according to claim 1 or 2, characterised by the fact that the wire-frame vertices, corresponding to the model feature points, on the basis of which facial animation parameters are determined in the analytic phase, are identified and said viseme interpolation operations are conducted by applying transforms on feature points for each viseme, for animating a wire-frame based model.
4. Method according to claim 3, characterised by the fact that, for each position to be assumed by the model in said synthesis phase, the transforms are applied only to the vertices of the wire-frame corresponding to the feature points and the transforms are extended to the remaining vertices by means of a convex combination of the transforms applied to the vertices of the wire-frame corresponding to the feature points.
5. Method according to claim 1, characterised by the fact that said visemes are converted into co-ordinates of the feature points of the face of the speaker, followed by conversion of said co-ordinates into said low-level facial animation parameters, as described in standard ISO/IEC 14496.
6. Method according to claim 5, characterised by the fact that said low-level facial animation parameters, representing the co-ordinates of feature points, are obtained by analysing the movements of a set of markers (7) which identify the feature points themselves.
7. Method according to claim 6, characterised by the fact that the data representing the co-ordinates of the feature points of the face are normalised according to the following method:
- a sub-set of markers are associated to a stiff object (8) applied to the forehead of the speaker;
 - the face of the speaker is set, at the beginning of the recording, to assume a position corresponding as far as possible to the position of a neutral face model, as defined in standard ISO/IEC 14496, and a first frame of the face in such neutral position is obtained;
- for all frames subsequent to the first frame, the sets of co-ordinates are rotated and translated so that the co-ordinates corresponding to the markers of said sub-set coincide with the co-ordinates of the markers of the same sub-set in the first

frame.

add B1